The Bari Manifesto : An interoperability framework for essential biodiversity variables

Hardisty, Alex R.

2019-01

Hardisty , A R , Michener , W K , Agosti , D , Alonso Garcia , E , Bastin , L , Belbin , L , Bowser , A , Buttigieg , P L , Canhos , D A L , Egloff , W , De Giovanni , R , Figueira , R , Groom , Q , Guralnick , R P , Hobern , D , Hugo , W , Koureas , D , Ji , L , Los , W , Manuel , J , Manset , D , Poelen , J , Saarenmaa , H , Schigel , D , Uhlir , P F & Kissling , W D 2019 , ' The Bari Manifesto : An interoperability framework for essential biodiversity variables ' , Ecological Informatics , vol. 49 , pp. 22-31 . https://doi.org/10.1016/j.ecoinf.2018.11.003

http://hdl.handle.net/10138/298536
https://doi.org/10.1016/j.ecoinf.2018.11.003

Ecological Informatics 49 (2019) 22–31

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

The Bari Manifesto: An interoperability framework for essential biodiversity variables

Alex R. Hardisty[a,*], William K. Michener[b], Donat Agosti[c], Enrique Alonso García[d], Lucy Bastin[e], Lee Belbin[f], Anne Bowser[g], Pier Luigi Buttigieg[h], Dora A.L. Canhos[i], Willi Egloff[c], Renato De Giovanni[i], Rui Figueira[j,w], Quentin Groom[k], Robert P. Guralnick[l], Donald Hobern[m], Wim Hugo[n], Dimitris Koureas[o], Liqiang Ji[p], Wouter Los[q], Jeffrey Manuel[r], David Manset[s], Jorrit Poelen[t], Hannu Saarenmaa[u], Dmitry Schigel[m], Paul F. Uhlir[v], W. Daniel Kissling[q]

[a] School of Computer Science & Informatics, Cardiff University, Queens Buildings, 5 The Parade, Cardiff CF24 3AA, United Kingdom
[b] College of University Libraries & Learning Sciences, MSC04 2815, University of New Mexico, NM 87131-0001, USA
[c] Plazi, Zinggstrasse 16, 3007 Bern, Switzerland
[d] Franklin Institute, University of Alcalá, Madrid, Spain
[e] European Commission, Joint Research Centre (JRC), Directorate D - Sustainable Resources, Knowledge Management Unit, Via Enrico Fermi 2749, I-21027 Ispra, VA, Italy
[f] The Atlas of Living Australia, PO Box 1700, Canberra, ACT. 2601, Australia
[g] Woodrow Wilson International Center for Scholars, 1300 Pennsylvania Ave., Washington, DC, USA
[h] HGF-MPG Group for Deep Sea Ecology and Technology, Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Am Handelshafen 12, 27570 Bremerhaven, Germany
[i] Centro de Referência em Informação Ambiental - CRIA, Campinas, São Paulo, Brazil
[j] CIBIO/InBIO-Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, Rua Padre Armando Quintas, 4485–601 Vairão, Portugal
[k] Botanic Garden Meise, Domein van Bouchout, B-1860 Meise, Belgium
[l] University of Florida Museum of Natural History, University of Florida at Gainesville, 358 Dickinson Hall, Gainesville, FL 32611-2710, USA
[m] GBIF Secretariat, Universitetsparken 15, 2100 København Ø, Denmark
[n] South African Environmental Observation Network, Cape Town, South Africa
[o] Naturalis Biodiversity Center, 2300RA Leiden, The Netherlands
[p] Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang, Beijing, 100101, China
[q] Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands
[r] South African National Biodiversity Institute, Pretoria, South Africa
[s] Be-Studys, Route de Meyrin 123, 1219 Châtelaine, Geneva, Switzerland
[t] 400 Perkins Street Apt 104, Oakland, CA 94610, United States
[u] University of Helsinki, Department of Forest Sciences, 00014 Helsinki, Finland
[v] P.O. Box 305, Callicoon, NY 12723, United States
[w] CEABN/InBIO-Centro de Estudos Ambientais 'Prof. Baeta Neves', Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisboa, Portugal

ARTICLE INFO

Keywords:
Essential biodiversity variables
Cyberinfrastructure
E-infrastructure
Data products
Informatics
Interoperability

ABSTRACT

Essential Biodiversity Variables (EBV) are fundamental variables that can be used for assessing biodiversity change over time, for determining adherence to biodiversity policy, for monitoring progress towards sustainable development goals, and for tracking biodiversity responses to disturbances and management interventions. Data from observations or models that provide measured or estimated EBV values, which we refer to as EBV data products, can help to capture the above processes and trends and can serve as a coherent framework for documenting trends in biodiversity. Using primary biodiversity records and other raw data as sources to produce EBV data products depends on cooperation and interoperability among multiple stakeholders, including those collecting and mobilising data for EBVs and those producing, publishing and preserving EBV data products. Here, we encapsulate ten principles for the current best practice in EBV-focused biodiversity informatics as 'The Bari Manifesto', serving as implementation guidelines for data and research infrastructure providers to support the emerging EBV operational framework based on trans-national and cross-infrastructure scientific workflows. The principles provide guidance on how to contribute towards the production of EBV data products that are

* Corresponding author at: School of Computer Science & Informatics, Cardiff University, Queens Buildings, 5 The Parade, Cardiff CF24 3AA, United Kingdom.
E-mail address: hardistyar@cardiff.ac.uk (A.R. Hardisty).

https://doi.org/10.1016/j.ecoinf.2018.11.003
Received 30 July 2018; Received in revised form 7 November 2018; Accepted 15 November 2018
Available online 17 November 2018

globally oriented, while remaining appropriate to the producer's own mission, vision and goals. These ten principles cover: data management planning; data structure; metadata; services; data quality; workflows; provenance; ontologies/vocabularies; data preservation; and accessibility. For each principle, desired outcomes and goals have been formulated. Some specific actions related to fulfilling the Bari Manifesto principles are highlighted in the context of each of four groups of organizations contributing to enabling data interoperability - data standards bodies, research data infrastructures, the pertinent research communities, and funders. The Bari Manifesto provides a roadmap enabling support for routine generation of EBV data products, and increases the likelihood of success for a global EBV framework.

1. Introduction

Reducing and reversing the rate of biodiversity loss and averting harmful biodiversity change are accepted international goals. However, there is still no global, harmonized observation or data exchange system for delivering regular, timely, and readily comparable information on biodiversity change (Navarro et al., 2017). Changes in biodiversity can be measured in different dimensions and across multiple scales, such as genetic, taxonomic and trait diversity across ecological units (communities, populations, species, clades), as well as at the ecosystem and biome level and on different temporal and spatial scales.

A key mechanism for studying and reporting on biodiversity and its change across the different dimensions is the concept of Essential Biodiversity Variables (EBVs) (Pereira et al., 2013). These are a candidate set of 22 variables considered critical to representing different dimensions of biodiversity change. They cover genetic composition, species populations, species traits, community composition, ecosystem function, and ecosystem structure. Raw data and biodiversity measurements collected and harmonized over space and time, supplemented with modelled estimates where interpolation/extrapolation is needed, provide the necessary data basis for EBVs, allowing interpretation into high-level indicator information for assessing biodiversity change. This is especially the case when such data sets are assembled at fine scale and broad extent. These data sets and indicators derived from them can be used to measure achievement of policy goals such as the Aichi Biodiversity Targets set by the Convention on Biological Diversity (CBD, 2018a), or the United Nations Sustainable Development Goals (UN, 2018) and the national targets defined in National Biodiversity Strategies and Action Plans (NBSAP) (CBD, 2018b). They can also serve to define biodiversity management policies and priorities from local to global scale.

GEO BON (GEO BON, 2018a) is the part of the global Group on Earth Observations (GEO) (GEO, 2018) that works to improve acquisition and delivery of biodiversity observations and related services to decision makers and the scientific community. EBVs are being defined by GEO BON to support biodiversity observation networks worldwide that contribute data to underpin effective management policies for the world's biodiversity and ecosystem services (Navarro et al., 2017). The EBV approach has been further explored by biodiversity scientists, global infrastructure operators and legal and policy experts in the EU-funded GLOBIS-B project “Global Infrastructures for Supporting Biodiversity research” (Kissling et al., 2015). This project examined infrastructure services underpinning the EBV concept and how international cooperation among data and research infrastructure organizations – hereinafter referred to as ‘Biodiversity Research Infrastructures’ (BRIs) – can support EBV definition and development, the development of workflows that adequately capture and organise EBV measurements, and subsequent management of that data. This cooperation has discussed, for EBV classes such as species populations (Kissling et al., 2018a) and species traits (Kissling et al., 2018b), how, in a computer assisted environment harmonizing data collection and preparation, technical data management and workflow processes can lead to standardized and reproducible data products with common characteristics. From those discussions, it has become clear that making EBVs operational requires a globally interoperable, trans-national information systems framework with local to global extent.

The present article makes clear the nature of EBV data products and the role of BRIs in supporting these as standardized products. To begin, section 2 explains the role of EBVs in a value chain from primary observation data to EBV data products to synthesised indicators of biodiversity change. It posits the need for a trusted, dependable and stable body of EBV data products, maintained over time. Section 3 discusses the general steps and actions that are required to construct EBV data products. Section 4 provides an overview of a real-world case study designed to demonstrate our current capacity to create EBV data products and subsequent indicators that can be used for policy, and concludes with strategic recommendations for next steps based on the case study. Section 5 examines a variety of existing infrastructures, the services they presently offer, and how these infrastructures can contribute to the collection of primary data, processing data and constructing EBV data products, and publishing and preserving the final EBV data products. Operationalizing the efficient production of EBV data products depends upon the ability of existing infrastructures to cooperate and effectively coordinate their activities (Kissling et al., 2015). Section 6 describes many of the technical (both syntactic and semantic) and legal challenges that must be resolved to enable interoperability among existing infrastructures. Section 7 proposes ten principles aligned to best current practices that outline how BRIs can promote interoperability and more effectively contribute to the production of global EBV data products. These principles are named ‘The

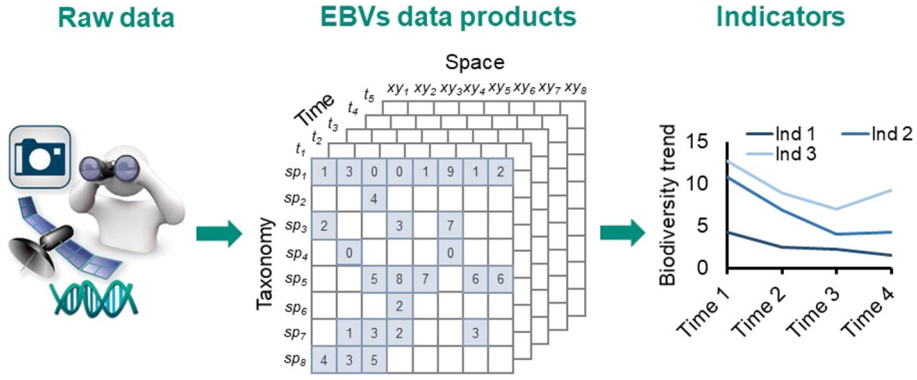

Fig. 1. Essential Biodiversity Variables (EBVs) are derived from raw data (i.e., primary observations) obtained, for example, from camera traps, field surveys, satellite remote sensing, and DNA sequencing. Harmonized, standardized and organised as packaged *EBV data products*, they provide the building blocks for indicator development. EBV data products can be conceptualized as cubes with dimensions of time, space and biology (taxonomy for example). Modified figure from (Kissling et al., 2018a).

Bari Manifesto', after the location (Bari, Italy) where they were agreed by BRIs representatives in February 2018. Section 8 highlights actions that can be taken to achieve better data interoperability by standards bodies, biodiversity research infrastructures, research communities, and research and infrastructure funders. Finally, section 9 offers conclusions about the future use of the Bari Manifesto.

2. EBVs as a fundamental resource for evaluating biodiversity trends

Since policy questions and management needs vary over time and political boundaries (e.g., between countries, regions, organizations), indicators of biodiversity change — such as those developed for monitoring progress towards targets set by the Convention on Biological Diversity (Butchart et al., 2010) — may also vary over space and time. A stable body of data corresponding to measured (and sometimes estimated) values of biodiversity that are comparable over space and time is critical to generate indicators (Pereira et al., 2013; Navarro et al., 2017) and, as such can have an extraordinarily positive impact on humankind's ability to address many of today's grand biodiversity and environmental challenges, such as quantifying, monitoring and mapping the loss of biodiversity and the corresponding loss or degradation of ecosystem services, the spread of invasive species and associated devastation of agricultural crops, and the spread of vector-borne diseases that have massive impacts on humans and livestock.

When organised and presented as discrete, defined packages of prepared and quality assured data, such data (here referred to as *EBV data products*) representing the whole body of data for multiple EBVs can be regarded as a stable intermediate data layer between *raw data* and varying *indicators* (Fig. 1).

Moreover, EBV data products that are sufficiently large (e.g., in terms of data volume, coverage, granularity) and comprehensive (in terms of temporal and spatial scales) would facilitate forecasting and assessing the impact of management interventions on biodiversity from national to global scales (Walters and Scholes, 2017).

The challenge is to agree on how to build such a dependable and stable body of sufficiently comprehensive data, and how to package and deliver it in a manner that can be most easily used to facilitate assessment and forecasting. Such agreement must be based upon cooperation, practicality and interoperability among those collecting and mobilising data with EBV potential, those processing, modelling and organising data, and those publishing and preserving data (Kissling et al., 2015). This can be compared with the situation currently prevailing for climate data, where stable, dependable essential climate variable (ECV) data are coming from the Global Observing System for Climate (GCOS) (GCOS, 2018; Bojinski et al., 2014).

3. Building EBV data products

Data products can be defined as a collected subset of one or more organization's data assets that are designed, packaged and presented to help a user solve a specific problem. An EBV data product is therefore a collection of data that offers standardized and comparable measurements and/or modelled estimates of the value an EBV takes at specific times and places. Hence, EBV data products will normally have components of geography and time (i.e., data for an area from multiple times), as well as one or more biological components such as taxonomy (Fig. 1). They can be built from multiple sources of raw data such as in situ monitoring, citizen science observations, genomic-based techniques and satellite/airborne remote sensing (Kissling et al., 2018a; Navarro et al., 2017; Walters and Scholes, 2017). And they are intended to be regularly updated and delivered to users over an extended period. Ideally, it should be possible to deliver EBV data products:

- For a stated geographic area;
- At defined spatial and temporal resolutions;
- For species, assemblages, ecosystems, or biomes of interest;
- With data held by relevant data repositories; and,
- By experts able to deploy the conceptual and operational framework of EBVs.

To build EBV data products requires the key dimensions (space, time, taxonomy, etc.), the attributes, and the acceptable uncertainties of raw data that can be usable for EBV purposes to be defined (Kissling et al., 2018a). Measurements must be in the desired format. They should be collected and processed following standardized protocols, and have sufficient associated metadata (Kissling et al., 2018a, 2018b). Data need to be consistently quality assured, using standard tests and associated assertions (Chapman et al., 2017; TDWG BDQ, 2018; Veiga et al., 2017). EBV data products should also meet minimum requirement standards for structure, packaging and metadata description. Such minimum standards have not yet been specified in the EBV context.

Workflows for generating EBV data products must cover all aspects of transforming raw data into published data products. This includes harmonizing data, and modelling where needed, as well as covering publishing and preserving the data product after it has been created (Kissling et al., 2018a, 2018b). From the view of BRIs, workflows should be independent of the underlying computational and data management infrastructure so that they are portable and adoptable across a wide range of possible infrastructure constructs. Raw data, the workflows and software should be traceable, allowing provenance to be tracked. EBV data production should be repeatable to allow easy updates as new data is collected. These needs can be met by using non-proprietary workflow formats, based for example, on the Common Workflow Language (CWL) (Amstutz et al., 2016), and standard provenance mechanisms (such as the W3C PROV family of specifications (Missier et al., 2013)). Resulting data products, including any component sub-parts, must be consistently structured (dimensioned, formatted, represented, packaged) and clearly described by metadata. They must be identifiable when published so they are discoverable and citable. Each data product must be preserved for the long-term as part of the dependable and stable body of EBV data. Much work remains to be done to achieve all this, keeping in mind that everything (raw data, data products, workflows, etc.) should be 'Findable, Accessible, Interoperable, Reusable', i.e., complying with the FAIR principles for scientific data management and stewardship (Wilkinson et al., 2016). This infers that both humans and machines can easily find, understand and exploit the data they need for their work.

4. Making EBV data products available for policy purposes: results from a case study

Generating EBV data products can require multiple BRIs to collaborate globally (Kissling et al., 2015), but the practical challenges regarding technical and legal interoperability have been little explored in the EBV context thus far. A recent case study (Hardisty et al., 2018) tested the ability of two mature infrastructures — the Atlas of Living Australia (ALA) and the Global Biodiversity Information Facility (GBIF) — to use a workflow approach (Atkinson et al., 2017; Hardisty et al., 2016) to deliver a species distribution EBV data product that could be used for evaluating the impact of three alien invasive species in Australia. This work revealed that workflow steps to discover, filter, and retrieve data were achievable within the capabilities of the two infrastructures, but that external tools, third-party sources and expert judgement were further needed to filter, process, check and merge the species distribution records into a prepared data product. The case study showed that workflows hold significant promise for delivering precise and maintainable data products (especially in terms of error prevention, automation and cost-reduction), but that further attention is needed in terms of automated processing and data integration. For instance, the standardization of data exchange structures, data access restrictions, and the right balance between applied human expertise

and machine automation are all issues that are recognized by BRIs to be further improved. Another important area where work is needed is on agreeing upon compact data/file structures for EBV data products, and how they can be handled by a wide-range of existing and to-be-developed software tools and services.

Moving from limited, experimental, proof-of-concept type studies such as the case study mentioned above to producing EBV data products for largely anonymous scientific and policy users is a key step for EBV development. While representative trials with real users are critical, organizations must also move towards robust and scalable solutions that provide a large-scale implementable framework for GEO BON across a wide range of EBV classes. This move, from prototypes to a production quality, factory-scale initiative requires clarity and alignment among multiple stakeholders on several strategic matters for EBV development (Table 1). Hence, not only technical issues need to be resolved, but also social and institutional issues across multiple bodies and BRIs. Scientists, infrastructure providers, informaticians, GEO BON working groups and policy end-users must therefore jointly identify what is feasible and useful. This work is beginning in the working groups and task forces of GEO BON (Navarro et al., 2017), but much remains to be addressed.

5. Roles for infrastructures in supporting EBVs

Informatics-based cyberinfrastructures/e-infrastructures currently support biodiversity science and ecology by collecting and providing primary data, aggregating or federating data for data discovery, integrating data, providing analysis and visualization services, and preserving data. Many BRIs offer multiple services (Table 2). Many infrastructures, such as the Atlas of Living Australia, the National Ecological Observatory Network, and the TRY Plant Traits Database serve as data providers or publishers. Map of Life is an integrator of information from several sources, assembling and modelling species range information and species lists for chosen geographic areas and producing summary indicators. Biodiversity Heritage Library is an integrating entry point to a network of institutions cooperating to digitize the legacy literature of biodiversity held in their collections and to make that available online. Others acts as aggregators of multiple sources. VertNET aggregates from natural science collections. GBIF aggregates not only from natural science collections but also from a wide range of other field-based, remote-sensed, genomic and literature sources. Some infrastructures, like DataONE and Catalogue of Life play a role more akin to federation, acting to bring participants closer together. DataONE federates through its member and coordinating nodes, offering centralised catalogues to distributed data repositories that can be independently accessed. Hundreds of other infrastructures serve as data repositories that integrate data, support discovery of data and provide delivery of data.

In the context of supporting EBVs, three roles of BRIs are particularly important: i) collecting and mobilising raw data with EBV potential; ii) processing, modelling and organising data into data products; and iii) publishing and preserving EBV data products. For the first role, existing data providers (including natural science collections), aggregators and others invest significant effort in mobilising raw data and making them openly available. For the second and third roles, a few infrastructures like the Atlas of Living Australia and GBIF have some limited capability, but in general they are not yet set up to process and organise data into EBV data products and to publish and preserve such products. This is not only due to missing consensus agreements on the actual work of producing EBV data products but also due to the high level of interoperability required among BRIs to underpin global generation of EBV data products once such agreements exist (Kissling et al., 2015). Moreover, improved collaboration and interoperability is not only required for in-situ measurements, but also for satellite remote sensing data where pathways of communication between the biodiversity community and the civilian space agencies (NASA, ESA) need to be improved (Leidner et al., 2017).

It is unlikely that new organizations will be created to generate EBV data products that will support national, regional and global research, conservation (e.g., parks, refuges), management and policy needs. Financial considerations aside, existing and cooperating BRIs could take on this role if data and infrastructure interoperability requirements can be addressed. Importantly, end-users and other stakeholders must be involved in defining EBV data products and the operational procedures needed for their production.

6. Interoperability among biodiversity data and research infrastructures

Improved interoperability between BRIs has been recognized as an important step for generating global EBV data products (Kissling et al., 2015). Interoperability refers to the capacity of computers and software to exchange and make use of data and information. This includes syntactic interoperability where two or more systems use the same data formats and communication protocol(s), and semantic interoperability when data are transferred meaningfully in a way that allows the receiving system to correctly understand and use the data exchanged (Heiler, 1995). Within the EBV context, *cross-domain* interoperability (Sartipi and Dehmoobad, 2008) is also important, and is achieved when multiple organizations agree upon common policies, principles and procedures.

Interoperability among BRIs today is still rudimentary, being mainly limited to exchanging data in a common format (i.e., syntactic interoperability). Darwin Core (DwC) (TDWG, 2018; Wieczorek et al., 2012), Ecological Metadata Language (EML) (Fegraus et al., 2005; KNB, 2018), ISO 19115 (ISO, 2018), Content Standards for Digital Geospatial Metadata and Biological Data Profile (FGDC 2018) and Access to Biological Collections Data (ABCD) (Holetschek et al., 2012; TDWG, 2007) are predominant choices for data and metadata formats. The successful adoption of these has enabled data providers to publish data and metadata in standard forms and has allowed infrastructures such as

Table 1
Strategic matters for further EBV development.

Topic	Clarity and support needed	Potentially responsible bodies
Clarification of policy priorities	On required EBV data products, in terms of prioritized species, assemblages, ecosystems, biomes, areas, scales, etc.	NGOs, governments, international organizations
Statements on national or thematic policy priorities	On required indicators, informing which EBV data products are missing	National and regional authorities
Coordinated monitoring schemes for primary data collection/production	Biodiversity Observation Networks (BONs) around the world to contribute data	GEO BON, individual BONs
Proven processing methods	Designed, tested and scientifically validated computational workflows to process primary observations into various EBV data products	Scientific and informatics communities, and their associated organizations; BRIs; GEO BON
Cooperation of data and research infrastructures	Producing, publishing and curating processed (EBV) data products in required formats	BRIs and their governing bodies; standards groups such as Biodiversity Information Standards (TDWG)
Overcoming legal constraints	Accessing and reusing data and achieving workflow interoperability	BRIs and their governing bodies; Research Data Alliance

Table 2
Examples of key cyberinfrastructure/e-infrastructures supporting biodiversity science and ecology and the principal services currently provided: C - Collection and organization of data; D - Discovery and access via data aggregation or data federation; A - Analysis and/or visualization of data; P - Data preservation; T - Training and education.

Infrastructure	Principal services	URL
Atlas of Living Australia (ALA), and the community of related 'Living Atlases'	C,D,A,P	https://www.ala.org.au/https://living-atlases.gbif.org/
Biodiversity Committee, Chinese Academy of Sciences	C,D,P,T	http://www.cncdiversitas.cn/
Biodiversity Heritage Library (BHL)	C,D,A,P	https://www.biodiversitylibrary.org/
Catalogue of Life (CoL)	D	http://www.catalogueoflife.org/
Data Observation Network for Earth (DataONE)	D,P,T	https://www.dataone.org/
Encyclopedia of Life (EoL)	D,T	http://eol.org/
Environmental Data Initiative	D,P,T	https://environmentaldatainitiative.org/
Global Biodiversity Information Facility (GBIF)	D,A,P,T	https://www.gbif.org/
Global Biotic Interactions (GloBI)	D,A	https://www.globalbioticinteractions.org/
Integrated Digitized Biocollections (iDigBio)	D,P,T	https://www.idigbio.org/
LifeWatch	A,T	https://www.lifewatch.eu/
Map of Life (MoL)	D,A	http://mol.org/
National Ecological Observatory Network (NEON)	C,D,P,T	https://www.neonscience.org/
National Specimen Information Infrastructure	C,D,T	http://nsii.org.cn/
Sistema de Informação sobre a Biodiversidade Brasileira (SiBBr)	C,D,A,P,T	http://www.sibbr.gov.br/
South African National Biodiversity Institute (SANBI)	C,D,A,P,T	http://www.sanbi.org/
speciesLink	D	http://www.splink.org.br/
TRY Plant Database	C,D,P	https://www.try-db.org/
VertNET	D,T	http://vertnet.org/

DataONE, GBIF and VertNET to aggregate and federate content across providers.

Semantic interoperability depends mainly on systems being in possession of a shared, congruent understanding of the context in which data exists and is exchanged. Attaching formal meaning to data through a process of interpretation and representing this with controlled vocabularies and relevant ontologies (i.e., creating interpretable information) is key to achieving semantic interoperability (Stocker et al., 2018). This kind of interoperability does not prevail today among BRIs, however.

Adopting similar syntactic and semantic interoperability regimes across multiple organizations (i.e., data collectors, BRIs, and users) can help significantly to optimise EBV data product generation and use. Several elements of cross-domain interoperability are primarily within the domain of the various stakeholder organizations, and mainly involve increasing the 'FAIRness' of data associated with adopting the FAIR guiding principles (Mons et al., 2017; Wilkinson et al., 2016); specifically, good data stewardship and metadata practices, including assignment of identifiers, common formats and machine-actionable metadata.

Having legal access to data, workflows and software, and their legal use and reuse across the domain is another kind of interoperability. Legal interoperability can be achieved when the accumulated conditions of use for each and all the datasets are met, and when users can legally access and use each dataset without seeking authorization from data rights holders on a case-by-case basis. The ideal goal for legal interoperability is when datasets are positively identified as having no legal restrictions (RDA-CODATA Legal Interoperability Interest Group, 2016). In the context of EBVs, formal designation with a CC0 copyright waiver or an open CC-BY license has been recommended (Kissling et al., 2018a; Kissling et al., 2018b). Although a waiver of copyright through CC0 makes sharing and reuse much easier, it doesn't waive a moral right to acknowledgement and attribution, which is important in the scientific context. The CC-BY license explicitly requires acknowledgement and attribution.

Below, we outline 'The Bari Manifesto' as a means for defining interoperability objectives for supporting creation and management of EBVs data products.

7. Ten principles for EBV data products – 'The Bari Manifesto'

A manifesto approach allows experts to establish and agree upon directions for technical infrastructure without being prescriptive about how or when infrastructure providers can achieve it, making it easier for organizations to agree to and adopt the guiding principles. Each of the ten principles have been formulated and agreed upon during a workshop organised by the GLOBIS-B project (Kissling et al., 2015), held in Bari, Italy 26–28 February 2018. As such, they reflect a consensus on the next steps needed towards improved interoperability among BRIs. Each principle, 'P' is stated as a desirable outcome, followed by explanatory information that includes both short-term and aspirational goals. Achieving the short-term goals within a reasonable timeframe should not be beyond any of today's infrastructure organizations, whilst achievement of the more aspirational goals will necessarily take longer.

7.1. Data management plan

P1. Projects or organizations developing EBV data products should have comprehensive data management plans.

Components of a data management plan should include information about: data structures and packaging; data formats and standards; metadata standards and tools; workflows; provenance; data quality control and quality assurance; referenced vocabularies and ontologies; policies that will be adhered to, including legal conditions of use; and the resource requirements (i.e., people, systems, training, software and services, repositories, maintenance) to produce and curate an EBV data product and the datasets upon which it depends (Michener, 2015; Michener, 2018). Furthermore, the plan should identify the desired or anticipated period of support for the EBV data product, as well as how that support will be sustained, and which organizations or individuals will provide the support.

7.2. Data structure

P2. EBV data products should adhere to agreed common dimensions for all products (i.e., time, space, name/taxonomy (where applicable), etc.). All data products should be accommodated in a common framework of dimensions and conform with established standards for representation formats.

Each EBV class/variable is likely to have its own distinct data model that should be part of the overall conceptual data structure. Nevertheless, each data model is likely to share elements in common with data models of other EBV classes/variables and the aim should be to achieve commonality wherever possible and appropriate. Clear definition of these data models and their common elements will help to identify what vocabulary definitions and relations are needed (see P8 below). The use of standard content and schema standards (e.g., NetCDF (UCAR, 2018), JSON (ECMA International, 2017) and newer compact data structures, for example (Ladra et al., 2017), encourages interoperability with the widest possible range of processing and visualization tools. Adherence to widely accepted nomenclatural authorities and name aggregators such as the Catalogue of Life (COL, 2018) and Integrated Taxonomic Information System (ITIS, 2018) resolves the challenges associated with integrating ambiguous, non-standardized taxonomic names (Parr and Thessen, 2018).

7.3. Metadata

P3. EBV data products and the data from which they are generated should have associated human- and machine-readable metadata that are compliant with accepted community standards and sufficient for purposes of data discovery, access, fitness-for-purpose evaluation, citation, interpretation and use.

Accepted community metadata specifications include those from bodies such as: Biodiversity Information Standards (TDWG, 2018), Federal Geographic Data Committee (FGDC, 2018), International Organization for Standardization (ISO, 2018), Research Data Alliance (RDA, 2018), Open Geospatial Consortium (OGC, 2018) and The World Wide Web Consortium (W3C, 2018). The “Ecological Metadata Language” specification (KNB, 2018) and “Minimum Information about any (X) Sequence” (MIxS) specification (Yilmaz et al., 2011) are also relevant metadata specifications.

7.4. Data quality

P4. Each EBV data product and its component sub-parts should undergo quality assurance testing and include information sufficient to ascertain the quality assurance and quality control procedures employed, and to help determine whether the data are of sufficient quality to use for specific purposes.

Data quality decisions (Chapman, 2005) made during production of an EBV data product should be fully documented, including statements about criteria used and thresholds applied. Standard tests (e.g., TDWG BDQ, 2018) should be automated and implemented from data capture to aggregation. It is desirable that assertions resulting from data quality tests be available as standard annotations at the record level wherever appropriate. The generation of EBV data products can involve filtering based on record-level quality assertions. It can also involve automated aggregation of quality assertions to produce a quality evaluation at the product level. Report-back of quality assertions to data providers should promote corrections at the source.

7.5. Services

P5. EBV data products, component datasets, digital objects and other related services should expose their capabilities and be accessible through common, standardized Application Programming Interfaces.

Decomposing programmatic functionalities into discrete services and operations offered through standardized Application Programming Interfaces (APIs) makes it easier to implement, maintain and evolve services as needs change (Newman, 2015). Using the OpenAPI specification (OpenAPI, 2018), such services can present themselves identically across multiple infrastructures, even when underlying details of their implementation differ one from infrastructure to another. As a first step, the community should adopt existing research data management technologies for sharing and registration of EBV data products in catalogues, and for query and retrieval; for example, CKAN (CKAN, 2018), Dataverse (Dataverse, 2018) or DSpace (DSpace, 2018). The community should agree on standard configurations (profiles) for discovery and access to EBV data products. These services can later evolve to a broader range of community tools that cover processing, brokering, visualization and workflow execution.

7.6. Workflows

P6. Standard workflows for preparing, publishing and preserving EBV data products and the component datasets from which they are produced must be fully documented and published, thus allowing them to be replicated and executed elsewhere. Ideally, they should be documented in a non-proprietary manner.

Standard workflows are needed to ensure that data products are both reproducible and consistent over time (Liew et al., 2016, Atkinson et al., 2017). These workflows should be represented in a language such as Common Workflow Language (Amstutz et al., 2016) with the potential to be understood by different workflow management systems. This would contribute significantly to making them portable across underlying execution mechanisms in different infrastructures. In the context of EBVs, prototype workflows have been created for species distribution and abundance (Kissling et al., 2018a; Hardisty et al., 2018) and species traits (Kissling et al., 2018b), but these workflows need to be robustly implemented for concrete EBV data products. Furthermore, re-usable components of such workflows already exist, e.g., for occurrence data retrieval and taxonomic data cleaning and integration (Mathew et al., 2014), and for creating, applying, projecting and visualizing models for species distributions and range shifts (De Giovanni et al., 2016). Such existing components should be integrated into standard workflows for EBV data products.

7.7. Provenance

P7. It should be possible to trace the EBV production process from the product back to the raw data and to reproduce the process. Provenance information must be readable both by humans and by machines.

In the short-term, this principle implies that details of all elements used in production of the EBV data product, such as the raw measurement data, the software tools, and the workflows should be packaged together and preserved; for example, as a research object with a persistent identifier such as a Digital Object Identifier (DOI) (Belhajjame et al., 2015; Hugo et al., 2017). In the longer term, tools and libraries implementing the W3C PROV specifications (Missier et al., 2013), such as those listed under the openProvance initiative (openProvenance, 2018) should be employed throughout the production process to support automated provenance generation and tracking. This leads to the potential for provenance graphs to be automatically traversed to understand origins and dependencies.

7.8. Ontologies/vocabularies

P8. EBV data products should be described by standard, openly accessible and machine-readable vocabulary terms and conceptual relations (ontologies). These terms and relations should be presented in a simple way to promote wide usage.

An extensible ‘EBV application ontology’ covering the main components of an EBV semantic layer should be developed as an interoperable and complementary part of the Open Biological and Biomedical Ontology (OBO) Foundry of ontologies (OBO Foundry, 2018a; Smith et al., 2007). This ontology should, as far as possible, inherit from and coordinate with terms and concepts from existing sources such as those of Biodiversity Information Standards (TDWG) and the biodiversity science domain (e.g., Darwin Core (DwC)

(Wieczorek et al., 2012), Biological Collections Ontology (BCO) (OBO Foundry, 2018b; Walls et al., 2014), Environment Ontology (Buttigieg et al., 2013, 2016; ENVO, 2018), Population and Community Ontology (PCO) (OBO Foundry, 2018c; Walls et al., 2014), the OBO Foundry (OBO Foundry, 2018a; Smith et al., 2007) and the Semantic Web for Earth and Environmental Terminology (SWEET) collection (SWEET, 2018). Persistent efforts are required to converge or align the descriptions of primary data resources used in the production of EBV data products and to encourage the widespread adoption and use of vocabularies and ontologies by research communities because they are critical for successfully completing complex data integration tasks. OWL (W3C, 2012) or SPARQL (W3C, 2013) traversal and interpretation of metadata can be enabled by linking reference ontologies to metadata.

7.9. Data preservation

P9. EBV data products and associated underlying data should be preserved with an associated persistent identifier in a community supported, open and trusted repository.

Many community repositories exist, with well-known ones including Dryad, Figshare, and Zenodo. Many repositories relevant to the biodiversity and ecological sciences are catalogued in the Registry of Research Data Repositories (re3data.org, 2018). Trusted repositories are those certified by, for example CoreTrustSeal, Data Seal of Approval, or ICSU World Data System Certification to provide long-term, reliable and open access to digital data products, with most, if not all of them assigning persistent identifiers to their data holdings and meeting other well-defined criteria (CRL and OCLC, 2007; Stall et al., 2017).

7.10. Accessibility

P10. EBV data products must be timely, open and FAIR (Findable, Accessible, Interoperable and Reusable).

Data should be mobilised and processed from the point of production to ensure they are available in a timely manner for research and policy needs. There should not be undue delays or hindrances for reasons other than simply the time it takes to perform the procedures. Appropriate attribution should be given and the fewest possible limitations placed on use. EBV data and data products should, to the greatest extent possible be open for anyone to freely access, use, modify, and share for any purpose (Kissling et al., 2018a). Copyright waivers and licenses (if any) should be offered in both human- and machine-readable form.

The FAIR data principles (Wilkinson et al., 2016), besides providing the basics for determining legal interoperability, cover requirements relating to metadata, identification, cataloguing and licensing. The FAIR principles aim to assist humans and machines in their discovery of, access to, integration and analysis of task-appropriate scientific data and their associated algorithms and workflows. EBV data products and the workflows necessary to create and use them must be findable and accessible via standard persistent identifier resolution mechanisms (for example, Digital Object Identifiers (DOI)) (Hugo et al., 2017). Their metadata, including information on legal use conditions must be openly available, and searchable via a catalogue maintained by an acknowledged authority, for example, GEO BON.

8. Next steps in enabling data interoperability

The ten principles comprising the Bari Manifesto (Section 7) provide a roadmap for supporting the syntactic, semantic, cross-domain and basic legal interoperability capabilities (Section 6) required to enable routine generation of EBV data products. The guidance embodied in the principles increases the likelihood of success for a global EBV framework, and allows stakeholder organizations in developing EBV data products (e.g., data providers, IT infrastructures) to retain autonomy and flexibility in achieving interoperability goals in ways that are most appropriate to their own businesses. Creating the full spectrum of interoperability solutions is expensive, time-consuming, far outside the scope and purview of any one organization, and not without its difficulties. Solving these interoperability challenges requires resources, coordination and contributions from: i) data standards bodies; ii) research data infrastructures; iii) the pertinent research communities; and iv) research and infrastructure funders. Below, we highlight some of the specific actions related to fulfilling the Bari Manifesto principles (P1-P10) that can be taken by each of the four groups of organizations contributing to enabling data interoperability.

8.1. Standards bodies

Standards bodies have a central role in resolving domain-specific interoperability issues that are especially relevant to EBV data. For example, Biodiversity Information Standards (TDWG, 2018) relies on Interest and Task Groups to develop standards relating to biodiversity data. The OBO Foundry (OBO Foundry, 2018a) represents a collective of ontology developers from many domains that collaboratively contribute to developing interoperable, non-overlapping ontologies based on shared principles and exemplary ontology models. The Research Data Alliance (RDA, 2018) relies on global Working and Interest Groups to develop the social and technical infrastructure that facilitates and promotes open data sharing both within and across domains.

Efficient and seamless creation of EBV data products requires that increased and concerted attention be focused on identifying, creating or refining a small number of acceptable standards suitable for: i) formatting and packaging digital objects (P2); ii) documenting fitness-for-purpose and other information needed for interpretation and use (P3); iii) quality assurance testing and assertion (P4); iv) representing workflows (P6); v) tracing the provenance of data and algorithms (P7); vi) capturing and representing EBV vocabulary terms and conceptual relations (i.e. ontologies) (P8); and vii) clarifying the degree of accessibility (e.g., adherence to FAIR guidelines, etc.) (P10).

8.2. Biodiversity Research Infrastructures (BRIs)

BRIs (Table 2) support the biodiversity and ecological sciences by making primary data more discoverable, interpretable and usable via publication, aggregation, federation, and the provision of applications (e.g., analytical and visualization tools). Yet, the widespread generation and use of a corpus of EBV data products will require significant new resources and capacity building, ranging from adopting a common policy(ies) for building EBV data products, agreeing on an architecture for storing (preserving), publishing, to discovering and retrieving EBV data and products. Success also depends on national and international policy bodies providing guidance on priorities and removing legal and financial barriers to cooperation.

Although new capabilities are needed, existing BRIs are well-positioned to experiment with and test the suitability of alternative standards and protocols. BRIs singly or collaboratively can propose and test: i) pilot implementations of data product quality evaluations (P4); ii) standard services for discovering and accessing EBV data products and underlying data (P5); and iii) mechanisms for exposing workflows (P6), provenance traces (P7), ontologies and controlled vocabularies (P8), and accessibility information (P10). Ideally, the experiments and prototyping activities would be jointly planned and coordinated, possibly by an organization such as GEO BON with its BONs that has ties to a range of stakeholder communities (e.g., users, funders, infrastructure providers).

It is possible to imagine, for example, a scenario in which the 'Living Atlases' codebase (ALA Community, 2018; Lecoq et al., 2018) incorporates a capability to produce species-level EBV data products (especially species populations) that could become a standard for biodiversity information systems around the world. GBIF could hold the responsibility for publishing and preserving specific EBV data products,

with GEO BON's "BON-in-a-Box" initiative (GEO BON, 2018b) providing tools to support consistent collection of new data. Introducing a "Living Atlas" into a country delivers a national biodiversity information system compatible with EBV data product generation. Adopting BON-in-a-Box provides a set of components that can enable national targeted monitoring and data collection capability feeding new, prioritized data into that national information system. The built-in EBV capability provides the outputs to feed regional to global indicators.

8.3. Research communities

Research communities are characterised by their domain focus and their constituencies – individual researchers, professional societies, synthesis centres, and related entities. Such communities explicitly or implicitly establish their own community norms and are poised to identify research needs and the data and algorithms required to address specific questions and hypotheses in a scientific domain.

Research communities are essential contributors to: i) specifying the content and scale of the EBV data products, and agreeing upon how the raw data will be managed and processed (P1); ii) ascertaining and ensuring data fitness-for-use (P4); iv) developing the workflows steps necessary to create data products (P6); v) contributing to development of ontologies and controlled vocabularies (P8); vi) determining which community repositories to use (P9); and vii) making their data findable, accessible, interoperable and reusable (P10). Synthesis centres may be especially well-positioned to prototype different EBV data products and pioneer some of the major cultural changes and standardization activities necessary for widespread creation and use of EBV data products.

8.4. Research and infrastructure funders

Public and private funders have been instrumental in supporting research in the biodiversity and ecological sciences and in building the necessary research and information technology infrastructures. Funders also guide the evolution of research and technology by sponsoring new research initiatives and setting policies (e.g., requiring data management plans, data sharing, and research transparency).

Several of the manifesto principles involve exploitation of technologies and approaches that are underdeveloped in the present field. These would clearly benefit from substantive funding and new initiatives in research and skills development. In research, for example: work could be developed on: i) exploring the use of Digital Object Architecture (Kahn and Wilensky, 2006) for structuring and managing EBV-related assets (P2); ii) developing automated quality assurance procedures to be applied during creation of EBV data products (P4); iii) advancing workflow and provenance technologies (P6, P7); and iv) filling critical gaps in ontology development (P8). Additionally, given the broad spectrum of biodiversity data and all possible indicators, more specific case studies are essential to discover and validate the most effective and comprehensive ways of implementing EBV data products. Such initiatives should preferably cover the whole process from raw data to real indicators, trying to engage stakeholders and communities along the entire EBV value chain. This would also help to develop new - or agree upon existing - protocols and standards.

In skills development, it may be especially fruitful to emphasize the training of data scientists capable of operating in this field, from data custodians up to data curators, biodiversity informaticians and "big data" analytics experts (Demchenko et al., 2016; Wiktorski et al., 2017). Funders in conjunction with other key stakeholders can have a disproportionately positive influence on developing new policies and legislation that are effective for opening access to and sharing of data (for example, see ROARMAP, 2018 and FAIRsharing, 2018). This is especially so for supporting cross-border and cross-domain research, resource management, and decision-making.

9. Conclusions

Considerable progress has been made in understanding how to operationalise the EBV concept, on how BRIs can work together to implement procedures for constructing, publishing and preserving EBV data products and how both policy authorities and scientific communities can benefit from a dependable and stable body of EBV data products. A coordinated test on biodiversity change related to invasive alien species in Australia recently demonstrated that EBV data products are feasible in practice but significant challenges remain (Hardisty et al., 2018). 'The Bari Manifesto' has the potential to significantly improve the ability of biodiversity research infrastructures to support the EBV production process, and to bring about general improvements in data interoperability for biodiversity and ecological sciences.

Acknowledgments

The work reported in this article has been performed by the GLOBal Infrastructures for Supporting Biodiversity research (GLOBIS-B) project (www.globis-b.eu) funded by the European Union Horizon 2020 Programme, grant no. 654003 (2015 - 2018). W. Michener was supported by the US National Science Foundation (grant nos. 1430508, 1757207). W. Daniel Kissling acknowledges support from the University of Amsterdam Faculty Research Cluster 'Global Ecology'.

References

ALA Community, 2018. ALA Community Living Atlases. [WWW Document] URL. https://living-atlases.gbif.org/ (accessed 6.15.18).

Amstutz, P., Crusoe, M.R., Tijanić, N., Chapman, B., Chilton, J., Heuer, M., Kartashov, A., Leehr, D., Ménager, H., Nedeljkovich, M., Scales, M., Soiland-Reyes, S., Stojanovic, L., 2016. Common Workflow Language v1.0. https://doi.org/10.6084/m9.figshare.3115156.v2.

Atkinson, M., Gesing, S., Montagnat, J., Taylor, I., 2017. Scientific workflows: past, present and future. Futur. Gener. Comput. Syst. 75, 216–227. https://doi.org/10.1016/j.future.2017.05.041.

Belhajjame, K., Zhao, J., Garijo, D., Gamble, M., Hettne, K., Palma, R., Mina, E., Corcho, O., Gómez-Pérez, J.-M., Bechhofer, S., Klyne, G., Goble, C., 2015. Using a suite of ontologies for preserving workflow-centric research objects. Web Semant. Sci. Serv. Agents World Wide Web 32, 16–42. https://doi.org/10.1016/j.websem.2015.01.003.

Bojinski, S., Verstraete, M., Peterson, T.C., Richter, C., Simmons, A., Zemp, M., 2014. The concept of essential climate variables in support of climate research, applications, and policy. Bull. Am. Meteorol. Soc. 95 (9), 1431–1443. https://doi.org/10.1175/BAMS-D-13-00047.1.

Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., Carr, G.M., Chanson, J., Chenery, A.M., Csirke, J., Davidson, N.C., Dentener, F., Foster, M., Galli, A., Galloway, J.N., Genovesi, P., Gregory, R.D., Hockings, M., Kapos, V., Lamarque, J.-F., Leverington, F., Loh, J., McGeoch, M.A., McRae, L., Minasyan, A., Morcillo, M.H., Oldfield, T.E.E., Pauly, D., Quader, S., Revenga, C., Sauer, J.R., Skolnik, B., Spear, D., Stanwell-Smith, D., Stuart, S.N., Symes, A., Tierney, M., Tyrrell, T.D., Vié, J.-C., Watson, R., 2010 Apr 29. Global biodiversity: indicators of recent declines. Science 1187512. https://doi.org/10.1126/science.1187512.

Buttigieg, P., Morrison, N., Smith, B., Mungall, C.J., Lewis, S.E., 2013. The environment ontology: contextualising biological and biomedical entities. J. Biomed. Semantics 4, 43. https://doi.org/10.1186/2041-1480-4-43.

Buttigieg, P.L., Pafilis, E., Lewis, S.E., Schildhauer, M.P., Walls, R.L., Mungall, C.J., 2016. The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. J. Biomed. Semantics 7 (57). https://doi.org/10.1186/s13326-016-0097-6.

CBD, 2018a. Strategic Plan for Biodiversity 2011–2020, Including Aichi Biodiversity Targets. Convention on Biological Diversity. [WWW Document] URL. https://www.cbd.int/sp/default.shtml (accessed 6.6.18).

CBD, 2018b. National Biodiversity Strategies and Action Plans (NBSAPs). In: Convention on Biological Diversity, [WWW Document]. URL. https://www.cbd.int/nbsap/ (accessed 6.6.18).

Chapman, A.D., 2005. Principles of Data Quality. Global Biodiversity Information Facility. https://doi.org/10.15468/doc.jrgg-a190.

Chapman, A., Saraiva, A., Belbin, L., Veiga, A., Nicholls, M., Zermoglio, P., Morris, P., Schigel, D., Thompson, A., 2017. Fitness for use: the BDQIG aims for improved Stability and Consistency. Proc. TDWG 1, e20240. https://doi.org/10.3897/tdwgproceedings.1.20240.

CKAN, 2018. Comprehensive Knowledge Archive Network. [WWW Document]. URL. https://ckan.org/ (accessed 11.02.18).

COL, 2018. Catalogue of Life. WWW Document] URL. http://www.catalogueoflife.org/ (accessed 10.10.18).

CRL and OCLC, 2007. Trusted Repositories Audit & Certification: Criteria and Checklist,

Version 1.0. Online Computer Library CenterInc., Dublin, Ohio, and The Center for Research Libraries, Chicago [WWW Document]. URL. http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf (accessed 6.12.18).
Dataverse, 2018. Dataverse Open Source web application. [WWW Document]. URL. https://dataverse.org/ (accessed 11.02.18).
De Giovanni, R., Williams, A.R., Ernst, V.H., Kulawik, R., Fernandez, F.Q., Hardisty, A.R., 2016. ENM Components: a new set of web service-based workflow components for ecological niche modelling. Ecography 39, 376–383. https://doi.org/10.1111/ecog.01552.
Demchenko, Y., Belloum, A., Los, W., Wiktorski, T., Manieri, A., Brocks, H., Becker, J., Heutelbeck, D., Hemmje, M., Brewer, S., 2016. EDISON Data Science Framework: a Foundation for Building Data Science Profession for Research and Industry. In: 2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom). IEEE, pp. 620–626. https://doi.org/10.1109/CloudCom.2016.0107.
DSpace, 2018. DSpace Open Source Repository Software. [WWW Document]. URL. https://duraspace.org/dspace/ (accessed 11.02.18).
ECMA International, 2017. Standard ECMA-404: The JSON Data Interchange Syntax. [WWW Document]. URL. http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf (accessed 6.6.18).
ENVO, 2018. Environment Ontology. [WWW Document URLhttp://environmentontology.org/ ((accessed 6.6.18)).
FAIRsharing, 2018. A Curated, Informative and Educational Resource on Data and Metadata Standards, Inter-Related to Databases and Data Policies. [WWW Document]. URL. https://fairsharing.org/ (accessed 10.10.18).
Fegraus, E.H., Andelman, S., Jones, M.B., Schildhauer, M., 2005. Maximizing the Value of Ecological Data with Structured Metadata: an Introduction to Ecological Metadata Language (EML) and Principles for Metadata creation. Bull. Ecol. Soc. Am. https://doi.org/10.2307/bullecosociamer.86.3.158.
FGDC, 2018. Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata Technical Specification and Biological Data Profile of the Content Standard for Digital Geospatial Metadata. [WWW Document]. URL. https://www.fgdc.gov/metadata/csdgm-standard (accessed 10.10.18).
GCOS, 2018. Global Observing Systems Information Center (GOSCI). [WWW Document]. URL. https://www.ncdc.noaa.gov/gosic (accessed 10.26.18).
GEO, 2018. Group on Earth Observations. [WWW Document]. URL. https://www.earthobservations.org/ (accessed 6.6.18).
GEO BON, 2018a. Group on Earth Observations Biodiversity Observations Networks. WWW Document URL. https://geobon.org/ (accessed 6.6.18).
GEO BON, 2018b. GEO BON BON-in-a-Box (Biodiversity Observation Network in a Box). [WWW Document]. URL. https://boninabox.geobon.org/ (accessed 6.15.18).
Hardisty, A.R., Bacall, F., Beard, N., Balcázar-Vargas, M.-P., Balech, B., Barcza, Z., Bourlat, S.J., Giovanni, R., Jong, Y., Leo, F., Dobor, L., Donvito, G., Fellows, D., Guerra, A.F., Ferreira, N., Fetyukova, Y., Fosso, B., Giddy, J., Goble, C., Güntsch, A., Haines, R., Ernst, V.H., Hettling, H., Hidy, D., Horváth, F., Ittzés, D., Ittzés, P., Jones, A., Kottmann, R., Kulawik, R., Leidenberger, S., Lyytikäinen-Saarenmaa, P., Mathew, C., Morrison, N., Nenadic, A., Hidalga, A.N., Obst, M., Oostermeijer, G., Paymal, E., Pesole, G., Pinto, S., Poigné, A., Fernandez, F.Q., Santamaria, M., Saarenmaa, H., Sipos, G., Sylla, K.-H., Tähtinen, M., Vicario, S., Vos, R.A., Williams, A.R., Yilmaz, P., 2016. BioVeL: a virtual laboratory for data analysis and modelling in biodiversity science and ecology. BMC Ecol. 16. https://doi.org/10.1186/s12898-016-0103-y.
Hardisty, A.R., Belbin, L., Hobern, D., McGeoch, M.A., Pirzl, R., Williams, K.J., Kissling, W.D., 2018 November 15. Towards Essential Biodiversity Variables data products for monitoring alien invasive species. Environ. Res. Lett *Submitted; revision under review and acceptance awaited*.
Heiler, S., 1995. Semantic interoperability. ACM Comput. Surv. 27, 271–273. https://doi.org/10.1145/210376.210392.
Holetschek, J., Dröge, G., Güntsch, A., Berendsohn, W.G., 2012. The ABCD of primary biodiversity data access. Plant Biosyst. - An Int. J. Deal. with all Asp. Plant Biol. 146, 771–779. https://doi.org/10.1080/11263504.2012.740085.
Hugo, W., Hobern, D., Kõljalg, U., Tuama, É.Ó., Saarenmaa, H., 2017. Global infrastructures for biodiversity data and services. In: Walters, M., Scholes, R.J. (Eds.), The GEO Handbook on Biodiversity Observation Networks. Springer International Publishing, Cham, pp. 259–291. https://doi.org/10.1007/978-3-319-27288-7.
ISO, 2018. International Organization for Standardization. [WWW Document]. URL. https://www.ios.org/ (accessed 6.6.18).
ITIS, 2018. Integrated Taxonomic Information System. [WWW Document]. URL. https://www.itis.gov/ (accessed 10.10.18).
Kahn, R., Wilensky, R., 2006. A framework for distributed digital object services. Int. J. Digit. Libr. 6, 115–123. https://doi.org/10.1007/s00799-005-0128-x.
Kissling, W.D., Hardisty, A., García, E.A., Santamaria, M., De Leo, F., Pesole, G., Freyhof, J., Manset, D., Wissel, S., Konijn, J., Los, W., 2015. Towards global interoperability for supporting biodiversity research on essential biodiversity variables (EBVs). Biodiversity 16, 99–107. https://doi.org/10.1080/14888386.2015.1068709.
Kissling, W.D., Ahumada, J.A., Bowser, A., Fernandez, M., Fernández, N., García, E.A., Guralnick, R.P., Isaac, N.J.B., Kelling, S., Los, W., Mcrae, L., Mihoub, J.B., Obst, M., Santamaria, M., Skidmore, A.K., Williams, K.J., Agosti, D., Amariles, D., Arvanitidis, C., Bastin, L., De Leo, F., Egloff, W., Elith, J., Hobern, D., Martin, D., Pereira, H.M., Pesole, G., Peterseil, J., Saarenmaa, H., Schigel, D., Schmeller, D.S., Segata, N., Turak, E., Uhlir, P.F., Wee, B., Hardisty, A.R., 2018a. Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. Biol. Rev. 93, 600–625. https://doi.org/10.1111/brv.12359.
Kissling, W.D., Walls, R., Bowser, A., Jones, M.O., Kattge, J., Agosti, D., Amengual, J., Basset, A., van Bodegom, P.M., Cornelissen, J.H.C., Denny, E.G., Deudero, S., Egloff, W., Elmendorf, S.C., Alonso García, E., Jones, K.D., Jones, O.R., Lavorel, S., Lear, D., Navarro, L.M., Pawar, S., Pirzl, R., Rüger, N., Sal, S., Salguero-Gómez, R., Schigel, D., Schulz, K.-S., Skidmore, A., Guralnick, R.P., 2018b. Towards global data products of Essential Biodiversity Variables (EBVs) on species traits. Nat. Ecol. Evol. 2, 1531–1540. https://doi.org/10.1038/s41559-018-0667-3.
KNB, 2018. Knowledge Network for Biocomplexity: Ecological Metadata Language (EML). WWW Document URL. https://knb.ecoinformatics.org/#external//emlparser/docs/index.html (accessed 6.15.18).
Ladra, S., Paramá, J.R., Silva-Coira, F., 2017. Scalable and queryable compressed storage structure for raster data. Inf. Syst. 72, 179–204. https://doi.org/10.1016/J.IS.2017.10.007.
Lecoq, M.-E., Archambeau, A.-S., Cavière, F., Martin, D., dos Remedios, N., 2018. The living Atlases community in action: general introduction. Biodivers. Inf. Sci. Stand. 2, e2548. https://doi.org/10.3897/biss.2.25487.
Leidner, A.K., Skidmore, A.K., Turner, W.W., Geller, G.N., 2017. Essential Biodiversity Variables: a framework for communication between the biodiversity community and space agencies. AGU Fall Meeting Abstracts. http://adsabs.harvard.edu/abs/2017AGUFMGC11C0743L (accessed 7.30.18).
Liew, C.S., Atkinson, M.P., Galea, M., Ang, T.F., Martin, P., Van Hemert, J.I., 2016. Scientific Workflows: moving across Paradigms. ACM Comput. Surv. 49 (4), 66. https://doi.org/10.1145/3012429.
Mathew, C., Güntsch, A., Obst, M., Vicario, S., Haines, R., Williams, A., de Jong, Y., Goble, C., 2014. A semi-automated workflow for biodiversity data retrieval, cleaning, and quality control. Biodiversity Data Journal 2, e4221. https://doi.org/10.3897/BDJ.2.e4221.
Michener, W.K., 2015. Ten simple rules for creating a good data management plan. PLoS Comput. Biol. 11 (10), e1004525. https://doi.org/10.1371/journal.pcbi.1004525.
Michener, W.K., 2018. Project Data Management Planning. In: Recknagel, F., Michener, W. (Eds.), Ecological Informatics. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-59928-1_2.
Missier, P., Belhajjame, K., Cheney, J., 2013. The W3C PROV family of specifications for modelling provenance metadata. In: Proceedings of the 16th International Conference on Extending Database Technology - EDBT '13. ACM Press, New York, New York, pp. 773. https://doi.org/10.1145/2452376.2452478.
Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L.O.B., Wilkinson, M.D., 2017. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Inf. Serv. Use 37 (1), 49–56. https://doi.org/10.3233/ISU-170824.
Navarro, L.M., Fernández, N., Guerra, C., Guralnick, R., Kissling, W.D., Londoño, M.C., Muller-Karger, F., Turak, E., Balvanera, P., Costello, M.J., Delavaud, A., El Serafy, G., Ferrier, S., Geijzendorffer, I., Geller, G.N., Jetz, W., Kim, E.-S., Kim, H., Martin, C.S., McGeoch, M.A., Mwampamba, T.H., Nel, J.L., Nicholson, E., Pettorelli, N., Schaepman, M.E., Skidmore, A., Sousa Pinto, I., Vergara, S., Vihervaara, P., Xu, H., Yahara, T., Gill, M., Pereira, H.M., 2017. Monitoring biodiversity change through effective global coordination. Curr. Opin. Environ. Sustain. 29, 158–169. https://doi.org/10.1016/J.COSUST.2018.02.005.
Newman, S., 2015. Building microservices: Designing fine-grained systems. O'Reilly Media, Inc. ISBN 978–1–491-95035-7.
OBO Foundry, 2018a. Open Biological and Biomedical Ontology (OBO) Foundry. WWW Document URL. http://www.obofoundry.org/ (accessed 6.6.18).
OBO Foundry, 2018b. Biological Collections Ontology (BCO). WWW Document URL. http://www.obofoundry.org/ontology/bco.html (accessed 6.8.18).
OBO Foundry, 2018c. Population and Community Ontology (PCO). WWW Document URL. http://www.obofoundry.org/ontology/pco.html (accessed 6.8.18).
OGC, 2018. Open Geospatial Consortium. [WWW Document]. URL. http://www.opengeospatial.org/ (accessed 6.6.18).
OpenAPI, 2018. The OpenAPI Specification. WWW Document. URL. https://github.com/OAI/OpenAPI-Specification (accessed 11.02.18).
openProvenance, 2018. Open Provenance. [WWW Document]. URL. https://openprovenance.org/ (accessed 11.02.18).
Parr, C.S., Thessen, A.E., 2018. Biodiversity informatics. In: Recknagel, F., Michener, W.K. (Eds.), Ecological Informatics. Springer, Cham, Switzerland, pp. 375–399.
Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., Höft, R., Hurtt, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M., Wegmann, M., 2013. Essential biodiversity variables. Science 339, 277–278. https://doi.org/10.1126/science.1229931.
RDA, 2018. Research Data Alliance. [WWW Document]. URL. https://www.rd-alliance.org/ (accessed 6.6.18).
RDA-CODATA Legal Interoperability Interest Group, 2016. Legal Interoperability of Research Data: Principles and Implementation guidelines. https://doi.org/10.5281/ZENODO.162241.
re3data.org, 2018. re3data.org - Registry of Research Data Repositories [WWW Document]. URL 10.17616/R3D (accessed 6.12.18).
ROARMAP, 2018. Registry of Open Access Repository Mandates and Policies. [WWW Document]. URL. http://roarmap.eprints.org (accessed 10.10.18).
Sartipi, K., Dehmoobad, A., 2008. Cross-Domain Information and Service Interoperability, in: Proceedings of the 10th International Conference on Information Integration and Web-Based Applications & Services - IiWAS 08. ACM Press, USA, New York, pp. 25. https://doi.org/10.1145/1497308.1497318.
Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R.H., Shah, N., Whetzel, P.L., Lewis, S., Lewis, S., 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat. Biotechnol. 25, 1251–1255. https://doi.org/10.1038/nbt1346.
Stall, S., Robinson, E., Wyborn, L., Yarmey, L.R., Parsons, M.A., Lehnert, K., Cutcher-Gershenfeld, J., Nosek, B., Hanson, B., 2017. Enabling FAIR data across the Earth and

space sciences. Eos 98. https://doi.org/10.1029/2017EO088425.
Stocker, M., Paasonen, P., Fiebig, M., Zaidan, M.A., Hardisty, A., 2018. Curating Scientific Information in Knowledge Infrastructures. Data Science Journal 17, 21. https://doi.org/10.5334/dsj-2018-021.
SWEET, 2018. SWEET Overview. [WWW Document]. URL. https://sweet.jpl.nasa.gov/ (accessed 6.6.18).
TDWG, 2007. Access to Biological Collections Data task group. Access to Biological Collection Data (ABCD), Version 2.06. Biodiversity Information Standards (TDWG). [WWW Document]. URL. http://www.tdwg.org/standards/115 (accessed 6.15.18).
TDWG, 2018. Biodiversity Information Standards: Taxonomic Databases Working Group. [WWW Document]. URL. http://www.tdwg.org/ (accessed 6.6.18).
TDWG BDQ, 2018. Taxonomic Databases Working Group Biodiversity Data Quality (BDQ) Interest Group. [WWW Document]. URL. https://github.com/tdwg/bdq (accessed 6.6.18).
UCAR, 2018. Network Common Data Format (NetCDF). University Consortium for Atmospheric Research. [WWW Document]. URL. https://www.unidata.ucar.edu/software/netcdf/ (accessed 6.6.18).
UN, 2018. Sustain. Dev. Goals: 17 Goals to Transform Our World. United Nations [WWW Document]. URL. https://www.un.org/sustainabledevelopment/sustainable-development-goals/ (accessed 6.6.18).
Veiga, A.K., Saraiva, A.M., Chapman, A.D., Morris, P.J., Gendreau, C., Schigel, D., Robertson, T.J., 2017. A conceptual framework for quality assessment and management of biodiversity data. PLoS One 12, e0178731. https://doi.org/10.1371/journal.pone.0178731.
W3C, 2012. OWL 2 Web Ontology Language Document Overview (Second Edition) W3C Recommendation 11 December 2012. [WWW Document]. URL. https://www.w3.org/TR/owl2-overview/ (accessed 6.8.18).
W3C, 2013. SPARQL 1.1 Overview W3C Recommendation 21 March 2013. [WWW Document]. URL. https://www.w3.org/TR/sparql11-overview/ (accessed 6.8.18).
W3C, 2018. The World Wide Web Consortium. [WWW Document]. URL. https://www.w3.org/ (accessed 6.6.18).
Walls, R.L., Deck, J., Guralnick, R., Baskauf, S., Beaman, R., Blum, S., Bowers, S., Buttigieg, P.L., Davies, N., Endresen, D., Gandolfo, M.A., Hanner, R., Janning, A., Krishtalka, L., Matsunaga, A., Midford, P., Morrison, N., Tuama, É.Ó., Schildhauer, M., Smith, B., Stucky, B.J., Thomer, A., Wieczorek, J., Whitacre, J., Wooley, J., 2014. Semantics in support of Biodiversity Knowledge Discovery: an Introduction to the Biological Collections Ontology and Related Ontologies. PLoS One 9, e89606. https://doi.org/10.1371/journal.pone.0089606.
Walters, M., Scholes, R.J. (Eds.), 2017. The GEO Handbook on Biodiversity Observation Networks. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-27288-7.
Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., Vieglais, D., 2012. Darwin Core: an Evolving Community-developed Biodiversity Data Standard. PLoS One 7, e29715. https://doi.org/10.1371/journal.pone.0029715.
Wiktorski, T., Demchenko, Y., Belloum, A., 2017. Model Curricula for Data Science EDISON Data Science Framework. In: 2017 IEEE International Conference on Cloud Computing Technology and Science (CloudCom). IEEE, pp. 369–374. https://doi.org/10.1109/CloudCom.2017.60.
Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., Hoen, T., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3 (160018). https://doi.org/10.1038/sdata.2016.18.
Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J.R., Amaral-Zettler, L., Gilbert, J.A., Karsch-Mizrachi, I., Johnston, A., Cochrane, G., Vaughan, R., Hunter, C., Park, J., Morrison, N., Rocca-Serra, P., Sterk, P., Arumugam, M., Bailey, M., Baumgartner, L., Birren, B.W., Blaser, M.J., Bonazzi, V., Booth, T., Bork, P., Bushman, F.D., Buttigieg, P.L., Chain, P.S.G., Charlson, E., Costello, E.K., Huot-Creasy, H., Dawyndt, P., Desantis, T., Fierer, N., Fuhrman, J.A., Gallery, R.E., Gevers, D., Gibbs, R.A., Gil, I.S., Gonzalez, A., Gordon, J.I., Guralnick, R., Hankeln, W., Highlander, S., Hugenholtz, P., Jansson, J., Kau, A.L., Kelley, S.T., Kennedy, J., Knights, D., Koren, O., Kuczynski, J., Kyrpides, N., Larsen, R., Lauber, C.L., Legg, T., Ley, R.E., Lozupone, C.A., Ludwig, W., Lyons, D., Maguire, E., Methé, B.A., Meyer, F., Muegge, B., Nakielny, S., Nelson, K.E., Nemergut, D., Neufeld, J.D., Newbold, L.K., Oliver, A.E., Pace, N.R., Palanisamy, G., Peplies, J., Petrosino, J., Proctor, L., Pruesse, E., Quast, C., Raes, J., Ratnasingham, S., Ravel, J., Relman, D.A., Assunta-Sansone, S., Schloss, P.D., Schriml, L., Sinha, R., Smith, M.I., Sodergren, E., Spor, A., Stombaugh, J., Tiedje, J.M., Ward, D.V., Weinstock, G.M., Wendel, D., White, O., Whiteley, A., Wilke, A., Wortman, J.R., Yatsunenko, T., Glöckner, F.O., 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat. Biotechnol. 29, 415–420. https://doi.org/10.1038/nbt.1823.